

Matteo Riondato

| | | |
|-------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CONTACTS | Two Sigma Investments, LLC 100 Avenues of the Americas New York, NY 10013, USA | Voice: +1-646-292-6641 E-mail: matteo@twosigma.com WWW: http://cs.brown.edu/~matteo |
| RESEARCH INTERESTS | Data analytics, pattern mining, graph mining, randomized algorithms, social network analysis, algorithms for distributed/parallel architectures for big data, privacy issues in data analysis, statistical learning theory. | |
| CURRENT POSITION | Two Sigma Investments, LLC , New York, NY, USA Researcher, member of the Labs team, since July 2015 | |
| EDUCATION | Brown University , Providence, RI, USA Postdoctoral Research Associate, February 2015 – July 2015 <ul style="list-style-type: none">Supervisor: Prof. Eli Upfal Stanford University , Stanford, California, USA Postdoctoral Researcher, August 2014 – January 2015 <ul style="list-style-type: none">Supervisor: Prof. Chris Ré Brown University , Providence, Rhode Island, USA Ph.D., Computer Science, May 2014 <ul style="list-style-type: none">Dissertation: <i>Sampling-based randomized algorithms for Big Data analytics</i>Advisor: Prof. Eli UpfalBest student poster award at SIAM Int. Conf. on Data Mining (SDM'14), April 2014. M.Sc., Computer Science, May 2010 Università di Padova , Padua, Italy Laurea Magistrale (M.Sc.) 110/110 <i>cum laude</i> , Computer Engineering, July 2009 <ul style="list-style-type: none">Master Thesis: <i>Top-k frequent itemsets mining through sampling</i>Advisors: Prof. Andrea Pietracaprina, Prof. Eli Upfal, Fabio Vandin Laurea (B.Sc.), Information Engineering, July 2007 <ul style="list-style-type: none">Final Project: <i>Algorithmical foundations of cryptography</i>Advisor: Prof. Andrea Pietracaprina | |
| PUBLICATIONS IN JOURNALS | [3] M. Riondato and E. M. Kornaropoulos. Fast approximation of betweenness centrality through sampling. <i>Data Mining and Knowl. Disc.</i> , in press, 2015. [2] M. Riondato and E. Upfal. Efficient discovery of association rules and frequent itemsets through sampling with tight performance guarantees. <i>ACM Trans. on Knowl. Disc. from Data</i> , 8(4):20, 2014. [1] A. Pietracaprina, M. Riondato, E. Upfal, and F. Vandin. Mining top-k frequent itemsets through progressive sampling. <i>Data Mining and Knowl. Disc.</i> , 21(2):310–326 (special issue for ECML PKDD'10), 2010. | |
| PUBLICATIONS IN PEER-REVIEWED CONFERENCES | [13] VC-Dimension and Rademacher Averages: From Statistical Learning Theory to Sampling Algorithms – Tutorial Outline. To appear in <i>ACM SIGKDD Int. Conf. on Knowl. Disc. and Data Mining, KDD'15, Sidney, Australia, August 10 – 13, 2015</i> . | |
| PROCEEDINGS | [12] Mining Frequent Itemsets through Progressive Sampling with Rademacher Averages. To appear in <i>ACM SIGKDD Int. Conf. on Knowl. Disc. and Data Mining, KDD'15, Sidney, Australia, August 10 – 13, 2015</i> . [11] A. Anagnostopoulos, L. Becchetti, A. Fazzzone, I. Mele, and M. Riondato. The importance of being experts: Efficient max-finding in crowdsourcing. <i>ACM SIGMOD Int. Conf. on Management of Data, SIGMOD'15, Melbourne, Australia, May 31 – June 4, 2015</i> . [10] M. Riondato, D. García-Soriano, and F. Bonchi. Graph summarization with quality guarantees. <i>14th IEEE Int. Conf. on Data Mining, ICDM'14, Shenzhen, China, December 14–17, 2014</i> . [9] M. Riondato. Sampling-based data mining algorithms: Modern techniques and case studies. <i>European Conf. on Mach. Learn. and Knowl. Disc. in Databases, ECML PKDD'14, Nancy, France, September 15–19, 2014</i> . [8] M. Riondato and F. Vandin. Finding the true frequent itemsets. <i>14th SIAM Int. Conf. on Data Mining, SDM'14, Philadelphia, PA, USA, April 24–26, 2014</i> . | |

- [7] M. Riondato and E. M. Kornaropoulos. Fast approximation of betweenness centrality through sampling. *7th ACM Int. Conf. on Web Search and Data Mining, WSDM'14, New York, NY, USA, February 24–28, 2014*.
- [6] M. Riondato, J. A. DeBrabant, R. Fonseca, and E. Upfal. PARMA: A parallel randomized algorithm for association rules mining in MapReduce. *21st ACM Int. Conf. on Inform. and Knowl. Manag., CIKM'12, Maui, HI, USA, October 29 – November 02, 2012*.
- [5] M. Riondato and E. Upfal. Efficient discovery of association rules and frequent itemsets through sampling with tight performance guarantees. *European Conf. on Mach. Learn. and Knowl. Disc. in Databases, ECML PKDD'12, Bristol, UK, September 24–28, 2012*.
- [4] A. Pietracaprina, G. Pucci, M. Riondato, F. Silvestri, and E. Upfal. Space-round tradeoffs for MapReduce computations. *ACM Int. Conf. on Supercomputing, ICS'12, Venice, Italy, June 25–29, 2012*.
- [3] M. Akdere, U. Çetintemel, M. Riondato, E. Upfal, and S. B. Zdonik. Learning-based query performance modeling and prediction. *28th IEEE Int. Conf. on Data Engin., ICDE'12, April 1–5, 2012, Washington, DC, USA*.
- [2] M. Riondato, M. Akdere, U. Çetintemel, S. B. Zdonik, and E. Upfal. The VC-dimension of SQL queries and selectivity estimation through sampling. *European Conf. on Mach. Learn. and Knowl. Disc. in Databases, ECML PKDD'11, Athens, Greece, September 5–9, 2011*.
- [1] M. Akdere, U. Çetintemel, M. Riondato, E. Upfal, and S. B. Zdonik. The case for predictive database systems: Opportunities and challenges. *5th Biennial Conf. on Innovative Data Sys. Research, CIDR'11, Asilomar, CA, USA, January 9–12, 2011*.

| | | |
|--------|--------------------------------------------------------------------------------|------------------------------|
| AWARDS | SIAM/NSF Early Career Travel Award to SIAM SDM'15 | February 2015 |
| | SIAM Best Student Poster Award at SIAM SDM'14 | April 2014 |
| | SIAM Student Travel Grant to SIAM SDM'14 | April 2014 |
| | Brown University Dissertation Fellowship | Fall 2013 |
| | Yahoo! Research Summer Internship | Summer 2013 |
| | Research Fellowship from MIUR of Italy under Project AlgoDeep prot. 2008TFBWL4 | Summer 2011 |
| | Brown University Graduate Fellowship | Academic Year 2009–10 |

| | |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| INVITED TALKS | Efficient Frequent Itemsets mining through sampling <i>Database Research Group Seminar, University of Waterloo, 5/7/2014</i> . |
| | Efficient Frequent Itemsets mining through sampling <i>Database Group Seminar, MIT CSAIL, 04/10/2014</i> . |
| | Taming the challenges of Big Data with statistical data analytics. <i>Department of Computer Science, Boston College, 02/07/2014</i> . |
| | Fast betweenness estimation through sampling. <i>Data Management Group Seminar, Boston University, 10/17/2013</i> . |
| | Fast betweenness estimation through sampling. <i>Lab Research Seminar, Yahoo! Labs Barcelona, 6/13/2013</i> . |
| | Fast betweenness estimation through sampling. <i>Advanced Computing Group Talk, Department of Information Engineering, University of Padua, Italy, 5/30/2013</i> . |
| | Statistical learning theory meets knowledge discovery: Randomized algorithms for Big Data analytics. <i>Brown CS Industrial Partners Program Symposium, 4/25/2013</i> . |
| | Approximate aggregate database queries through sampling. <i>Invited Lecture for CSCI-2950-T, Brown University, 10/25/2011</i> . |
| | Graphs algorithms in MapReduce: Design choices and optimizations. <i>Invited Lecture for CSCI-2950-U, Brown University, 9/27/2011</i> . |
| | Statistical learning theory meets databases. <i>Advanced Computing Group Talk, Department of Information Engineering, University of Padua, Italy, 06/24/2011</i> . |
| | Top-k frequent itemsets mining through sampling. <i>Advanced Computing Group Talk, Department of Information Engineering, University of Padua, Italy, 9/16/2010</i> . |

| | | |
|--------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| ADDITIONAL RESEARCH EXPERIENCE | Yahoo! Labs , Barcelona, Spain <i>Summer Intern Research Scientist in the Web Mining Group</i> <ul style="list-style-type: none"> Worked with Dr. Francesco Bonchi and others on algorithms for graph summarization and for frequent itemsets mining from data streams in a distributed setting. | June – August 2013 |
| | Summer School on Massive Data Mining , Copenhagen, Denmark <i>Student</i> <ul style="list-style-type: none"> Attended the Summer School on Massive Data Mining organized by Prof. R. Pagh at IT University of Copenhagen. | August 8 – 10, 2012 |
| | Sapienza Università di Roma , Rome, Italy <i>Visiting Ph.D. Student</i> <ul style="list-style-type: none"> Worked with Prof. S. Leonardi, A. Anagnostopoulos, and L. Becchetti on algorithms for a model of crowdsourcing computation, and on algorithms for MapReduce. | June – July 2012 |
| | Università di Padova , Padua, Italy <i>Research Fellow</i> <ul style="list-style-type: none"> Worked with Prof. A. Pietracaprina, G. Pucci, and F. Silvestri on “The MapReduce Paradigm: computational model and algorithms”. Funded by MIUR of Italy under Project AlgoDEEP prot. 2008TFBWL4. | May – July 2011 |
| | Chalmers University of Technology , Gothenburg, Sweden <i>Visiting Ph.D. Student</i> <ul style="list-style-type: none"> Worked with Prof. Devdatt Dubhashi and helped organize a seminar on Probability and Computing. | August 2010 |
| TEACHING AND ADVISING TRAINING AND EXPERIENCE | Brown University , Providence, RI, USA <i>Visiting Grad Student</i> <ul style="list-style-type: none"> Worked with Prof. Eli Upfal on master thesis <i>Top-K Frequent Itemsets Mining through Sampling</i>. | October 2008 – June 2009 |
| | The Harriet W. Sheridan Center for Teaching and Learning (Brown University) , Providence, RI, USA <ul style="list-style-type: none"> Teaching Certificate I: Reflective Teaching, AY 2013–14 | |
| | Brown University , Providence, Rhode Island, USA <ul style="list-style-type: none"> Mentor, New Scientist Program Graduate-Undergraduate Mentoring Initiative, Spring 2014 Teaching Assistant, Probability and Computing, Spring 2012, Spring 2013, Spring 2014 Teaching Assistant, Probabilistic Methods in Computer Science, Fall 2010 | |
| | Program Committee member for the following conferences <ul style="list-style-type: none"> WWW’16, ACM CIKM’15, ACM KDD’15, ACM CIKM’14 | |
| | Reviewer for the following journals <ul style="list-style-type: none"> Data Mining and Knowl. Disc. (DMKD / DAMI), Knowl. and Info. Disc. (KAIS), IEEE Trans. Knowl. Disc. Eng. (TKDE), IEEE Trans. Netw. Sci. Eng. (TNSE), IEEE Trans. Parall. Distrib. Sys. (TPDS), IEEE Trans. Services Comput. (TSC), VLDB Journal | |
| SERVICE | Reviewer for the following conferences <ul style="list-style-type: none"> ISAAC’15, SIAM SDM’15, DISC’14, ACM WSDM’14, WWW’14, ICALP’14, IEEE BigData’13, MFCS’13, ACM WSDM’13, IEEE IPDPS’12, ACM ICS’12, RANDOM’11. | |
| | Brown University Graduate Student Council <i>President</i> <ul style="list-style-type: none"> Represented the interests of the entire graduate student community (approx. 2000 students) with the university administration and the broader community at all levels. Previously held positions include Vice-President of Administration (Jan. – April 2011) and representative for the Computer Science Department (Sept. 2009 – April 2011, Jan 2013 – May 2014). | April 2011 – December 2012 |
| | Brown University Graduate Council <i>Student Representative</i> <ul style="list-style-type: none"> Represented the graduate student community in the highest body governing graduate education. | September 2011 – May 2013 |
| | Brown University Strategic Planning Committee on Doctoral Education <i>Student Representative</i> <ul style="list-style-type: none"> Represented the graduate student community to develop the presidential strategic plan for the next decade. | September 2012 – May 2013 |
| | | |

Brown Computer Science Theory Lunch

Organizer

January – December 2010

- Responsible for organizing the weekly meeting of the Theory Group.

MEMBERSHIP

Association for Computing Machinery (ACM)

- Member of the Special Interest Group on Knowledge Discovery from Data (SIGKDD).

Institute of Electrical and Electronic Engineers (IEEE)

- Member of the IEEE Computer Society.

Society for Industrial and Applied Mathematics (SIAM)

- Member of the SIAM Activity Group on Data Mining and Analytics (SIAG/DMA).

SOFTWARE

CentrSampl

- Algorithm to estimate node betweenness centrality in large graphs. Based on [8].
<http://cs.brown.edu/~matteo/centrsampl.tar.bz2>.

PARMA

- Frequent itemsets and association rules mining algorithm for Hadoop MapReduce. Based on [7].
<http://cs.brown.edu/~matteo/parma.tar.bz2>.

FreeSBIE

Developer, Release Engineer for the 2.x series

April 2004 – July 2009

- Developed FreeSBIE, a Live-CD distribution of FreeBSD bootable from CD-ROM. Release Engineer for FreeSBIE 2.X. series, responsible for all aspects of the release. <http://www.FreeSBIE.org>.

The FreeBSD Project

src Committer

January 2006 – June 2013

- Contributed to the development of the FreeBSD UNIX operating system. Granted write access to the main source repository. Worked on the `jail` security feature and on handling and solving bug reports of various nature. Author of the *Jail* chapter in the FreeBSD Handbook. <http://www.FreeBSD.org>.

REFERENCES

Prof. Eli Upfal

Dept. of Computer Science

Brown University

eli@cs.brown.edu

Prof. Uğur Çetintemel

Dept. of Computer Science

Brown University

ugur@cs.brown.edu

Dr. Francesco Bonchi

Yahoo! Labs Barcelona

Yahoo!

bonchi@yahoo-inc.com

UP-TO-DATE VERSION

Available from http://cs.brown.edu/~matteo/matteo_riondato_cv.pdf